DOCUMENT RESUME

ED 329 595 TM 016 260

AUTHOR Abraham, Paul F.
TITLE The Reading Proficiency of Non-Native English Speaking Applicants to American Undergraduate Education: How Is It Assessed? How Should It Be Assessed? Qualifying Paper, Harvard University School of Education.
PUB DATE Nov 90
NOTE 58p.
PUB TYPE Dissertations/Theses - Practicum Papers (043) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *College Applicants; College Entrance Examinations; College Second Language Programs; Comparative Analysis; *English (Second Language); Higher Education; *Language Tests; Predictive Validity; *Reading Achievement; Reading Tests; Undergraduate Study
IDENTIFIERS *Test of English as a Foreign Language

ABSTRACT

English language assessment of non-native speakers of English (NNS) who apply for undergraduate study in the United States is examined. The analysis focuses on: the assessment of reading in light of its importance across the undergraduate curriculum, and the establishment of a base-line level of reading proficiency to which the proficiency of NNS can be compared. The paper addresses the following issues: the nature of college reading; current directions in reading assessment and the means by which these directions inform the assessment of applicants who are NNS; and the efficacy of the Test of English as a Foreign Language (TOEFL) as a measure of linguistic proficiency and college-level reading. The TOEFL is compared with other test instruments, and empirical evidence on the predictive validity of the TOEFL is reviewed. Historical perspectives are provided for both the college-level reading curriculum and the TOEFL. Insights from the developmental and skilled reading perspectives are also outlined. Reading theory and assessment, the direction of large-scale reading assessment, and skilled reading and assessment are discussed. The unique contributions of each of the three sections of the TOEFL to assessment of language ability are described. Recommendations relative to the use of the TOEFL that are pertinent to admissions officers, researchers in second language reading, and undergraduate faculty who teach NNS are offered. A 130-item list of references is included. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to: LE

In our judgment, this document is also of interest to the Clearinghouses noted to the right. Indexing should reflect their special points of view. HE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

PAUL ABRAHAM

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED329595

The Reading Proficiency of Non-Native English Speaking Applicants to American Undergraduate Education: How is it assessed? How should it be assessed?

Qualifying Paper
Submitted by
Paul F. Abraham

November 1990

TM016260

SECTION 1: COLLEGE-LEVEL READING 3

Historical Perspective 4

A. Changes in Curriculum

B. Changes in Student Body

Developmental Perspective 7

Skilled Reading Perspective 9

A. Skilled Reading

B. General Skills

C. Specific Tasks

Summary 15

SECTION 2: CURRENT DIRECTIONS IN READING ASSESSMENT

The assessment of college-level reading ability 17

A. *Reading Theory and Assessment* 17

B. *Where is large-scale reading assessment headed?* 19

C. *Skilled Reading and Assessment* 21

Assessment In ESL Reading 23

SECTION 3: THE TOEFL AS A READING/LANGUAGE MEASURE

A. What is the TOEFL?

Historical Context 25

Development 26

Description 27

B. What does the TOEFL measure?

Appropriateness for Non-native speakers 29

Comparisons with other measures

a. English Language Proficiency Tests 29

b. Aptitude and achievement tests 30

c. Non-standardized measures 30

Predictive Ability of the TOEFL 31

Unique Contributions from all three sections 32

C. How does the TOEFL measure reading comprehension? .. 34

Comparisions between the TOEFL, the Reading Subtest and various other tests 35

Content of the Reading Subtest of the TOEFL 36

Analysis of Three TOEFL Tests 38

A. TOEFL Vocabulary 38

B. TOEFL Reading 39

Summary 41

CONCLUSION 43

RECOMMENDATIONS 45

INTRODUCTION

Record numbers of non-native speakers of English (NNS) are entering American higher education (Zikopoulos, 1990). How can the English language proficiency of these NNS applicants be assessed to determine if these students are ready for the academic tasks which face them? This paper will examine the question of the English language assessment of (NNS) applicants for undergraduate study in the United States. The analysis will focus primarily on the assessment of reading in light of its importance across the undergraduate curriculum. Three major questions will guide the analysis: What is college reading? What are the current directions in reading assessment and how might these inform the assessment of NNS applicants? How well does the TOEFL, currently the most common language proficiency tool for NNS, measure the college-level reading of applicants to undergraduate education?

The number of non-native English speakers entering American higher education from abroad has been steadily increasing in recent years (Zikopoulos, 1990). Non-native English speakers are providing a rich source of enrollment at a period of continued decrease in the college-age population in the United States. These non-native speaking (NNS) applicants face the complex academic demands of higher education within a linguistic and cultural system not their own. Are NNS students, particularly at the undergraduate level, ready for the reading demands of American higher education? How can college and university admissions staff determine if these students are adequately prepared to read at the level required in undergraduate study? Standardized tests, such as the SAT, designed for NS (native speaking) students, do not adequately measure the linguistic ability of NNS students (Angelis, P.J., Swinton,S.S., Cowell, W.R., 1979). The Test of English as a Foreign Language (TOEFL), a standardized language test designed specifically for NNS students (Educational Testing Service),is the standard measure of linguistic proficiency for NNS university applicants. What does the TOEFL tell us about the reading ability of these students? What aspects of college-level reading does the TOEFL assess?

In section 1 of this paper, I will examine college-level reading theoretically and operationally to establish a level of reading proficiency from which to compare the proficiency of NNS students. In section 2, I will discuss current directions in first language reading assessment and show how these directions reflect our current knowledge about the reading process and skilled reading. I will also examine the types of reading assessments which have been used in research with non-native speakers of English. In section 3, I will examine the TOEFL (Test of English as a Foreign Language) as the primary language proficiency measure of NNS applicants for admission to American higher education. I will give an overview of the TOEFL, discuss how and what it measures with particular attention to its measurement of reading comprehension. I will then compare the assessment of reading on the TOEFL with current directions in L1 reading assessment for admission to American

higher education. I will give an overview of the TOEFL, discuss how and what it measures with particular attention to its measurement of reading comprehension. I will then compare the assessment of reading on the TOEFL with current directions in L1 reading assessment. Finally, I will make recommendations to admissions officers, second language researchers and undergraduate instructors about second language reading, the complexity of college-level reading, and the use of TOEFL scores. I will also suggest directions for future research to measure the reading proficiency of the NNS population.

The following questions will be addressed:

1. What is college-level reading? How is it defined theoretically and operationally? What skills and tasks are involved in college-level reading?

2. How is college-level reading assessed? What are the general directions in the assessment of reading in L1? How is reading assessed in ESL reading research?

3. What is the TOEFL and what does it measure? What aspects of the language proficiency does the TOEFL measure? Does the TOEFL measure college-level reading?

SECTION 1. COLLEGE-LEVEL READING: WHAT IS COLLEGE-LEVEL READING? WHAT SKILLS AND TASKS ARE INVOLVED?

Numerous recent publications (the *New York Times*, *The Wall Street Journal*, and the *Chronicle of Higher Education*) decry the lack of academic preparation among college students (Lederman, Ribaudo, and Ryzewic, 1985). Survey data from nearly one-half of the 2,800 institutions of higher education which admit first-year undergraduates nationwide corroborate perceptions found in the media: 85% of the responding institutions perceived poor academic preparation among freshmen to be a problem (Lederman, Ribaudo, and Ryzewic, 1985). Responses varied greatly across types of institutions (90% of two- year colleges found freshman preparation a problem whereas 45% of four year private universities and 62% of private four-year selective colleges perceived their preparation as a problem).

Increased access to college has greatly influenced the level of preparation among students. From 1966 to 1986, total college enrollment nearly doubled (from 6.3 million to 12.5 million) in great part due to the baby boom; the more recent enrollment increases have been largely due to increased access to higher education and primarily in 2-year colleges. More recently, nearly 50% of the college-age population is enrolled in college (Oakley, 1990). Because of the sheer numbers of current students and the great variance in their educational backgrounds, students' academic preparation is no longer at the same level as it was before open admissions. In what ways do current students lack preparation? What are the requirements of college-level academics?

Language is the first requirement in undergraduate education (Boyer, 1987), and educators would readily agree that the abilities to read and write effectively are the mainstays of this requirement. In fact, the process of writing has been the focus of extensive research in recent years in both first (Applebee, 1986; Flower & Hayes, 1981; Fulwiler, 1984; Hillocks, 1987)

and second language research (Raimes, 1985; Spack, 1988; Spack & Sadow, 1983; Zamel, 1982, 1983). The adoption of journal writing and writing "across the curriculum" have underscored the interrelatedness of writing and thinking. Many undergraduate instructors are now exploring writing as a process and a tool for learning rather than as a product or the demonstration of acquired knowledge. These written tasks in undergraduate courses are often based on student response to and understanding of source readings assigned within courses. Instructors assume competency in reading and expect that students are capable of reading material presented for summary, analysis, synthesis, or critique. What does this capability entail? What is college-level reading? In an attempt to provide a working definition, I will examine the construct of college level reading historically, developmentally and through a comparison to skilled reading.

College-level Reading: Historical Perspective

Reading problems have been addressed extensively in public education, but neglected at the college level according to Platt (1987). She criticizes the national report, "A Nation at Risk," which offers only one recommendation for higher education i.e. that admission standards be raised. Platt (1987) also reminds us (as does Heron, 1989) that under-preparation is not a recent phenomenon. Platt cites Maxwell's (1979) discussion of a 1915 report from 350 colleges to the U.S. Commissioner of Education, that there was "a gap between high school preparation and the expectations of college."

Current reading levels give cause for alarm. Plisko and Stern (1985) indicate that 28% of current college freshmen are identified as needing help in reading. Eurich & Kraetsch (1981) examined reading ability over time by comparing reading test scores of freshmen in 1928 with those of 1978 at the University of Minnesota. The findings indicated that the 1978 freshmen were reading, on average, one grade level below those of 1928. Community college students may be faring even less well than university students.

Beginning community college students are reading on average at the 8th or 9th grade level (Chall, 1985).

In analyzing the historical trends which have affected the nature of college-level reading from the 1600s to the present, Heron (1989) discusses two major trends: changes in curriculum affecting the scope and variety of college-level reading; and changes in the size and nature of the student body and in its purpose in studying.

A. Changes in Curriculum

How has college-level reading changed over time? Heron (1989) defines early college-level reading as involving "intensive reading focusing on classical and biblical studies," for factual interpretation and memorization. Heron cites Leedy (1959) who describes the content of early university curriculum as a "clearly defined body of knowledge about which students could learn and think" (p.24) emphasizing familiarity with a small number of texts and no outside reading. (Heron, citing Leedy (1959) reports that up until the mid-1800s, libraries were open only to seniors and faculty.) According to Heron, the purpose of the curriculum was to develop a Christian intellect and character based on a common set of truths. Only at the Master's level did students critically read and reason on paper.

Early changes in curriculum took place during the post Revolutionary period when college served as a place to learn about the new government (Heron). A new emphasis on English rather than the classics also occurred. Heron traces other changes to the onset of the Industrial and Scientific Revolutions when college became a place to learn things with increased specialization for industry and service. One would go to college to learn business, agriculture and politics (p.29). College reading became "wide-ranging and extensive, in many and varied domains" (p.17). The very nature of college-level reading changed radically and rapidly during the 1800s and these changes set the

stage for modern college reading [p.18, citing Leedy (1959) and Hofstadter and Smith (1961)].

Post Civil War brought more changes in college curricula as American academics emulated the comprehensiveness of German higher education (Silber, 1987) and broadened curricula. "Students had to assemble facts, wild with possibilities, questions and differing opinions" (Heron, p.34). Research became an important focus of the university. Students had to think scientifically. Answers were no longer known and students had to read to discover the relationship among facts. The curriculum was based on greater freedom and specialization, on inquiry, conflicting ideas and the demands of research (p.73).

This broadening of the curriculum has continued to the present. More recently, widespread diversification in the content of college courses and technological advances in information gathering and dissemination (from xeroxing to on-line computer searches) have broadened and amplified the reading demands of the current college curriculum.

B. Changes in the Student Body

Heron (1989) describes the typical student body in the early years of American higher education as small numbers of young men, similar in their pre-college preparation, studying for careers in the clergy. Silber (1987) concurs suggesting that Yale, Princeton and Columbia, founded before the Revolution, were "small institutions designed primarily for the education of ministers and magistrates" (p.38). Silber traces the changes in American higher education from 1900 when 2.3% of college-age Americans were enrolled to before World War II with 9.1% enrolled.

The post World War II increase in college-age population resulted in an unprecedented expansion of publicly funded higher education. The Civil

Rights movement of the 1960s also demanded accessibilty to higher education for the American underclass (See Cross, 1976). The current student body includes large numbers of men and women, of various ages and of widely diverse social and economic backgrounds, seeking advanced study in numerous areas (Heron, 1989). Current figures indicate that nearly 50% of college-age Americans are enrolled in some form of higher education (Oakley, 1990). This democratization of American higher education has not been welcomed by all and, for some, this "open enrollment" explains the under-preparation of students and the watering down of standards (See Silber, 1987). Still, others see American higher education as a great success and access to higher education as the hallmark of democracy (See Oakely, 1990).

College-Level Reading: Developmental Perspective

Developmental theories which view reading as a series of progressive stages over time offer another perspective on the nature of college-reading (Carroll, 1977; Chall, 1983).

Carroll (1977) sees reading as an on-going process intimately tied to the development of language and cognition. Carroll's theoretical framework suggests that the development of language affects the reading and cognition of readers at all levels of development. As a college reader's linguistic and cognitive experience broadens so too does his/her ability to understand text. Carroll's model helps illuminate the ties among the modes of learning available at the undergraduate level: lecture, discussion and reading.

Chall's (1983, p. 10) theory, which owes much to the work of Piaget [Inhelder and Piaget, 1958; Piaget, 1970], outlines reading development in six stages. These stages of reading, culminating in college-age and above, provide a specific description of reading at the college-level as well as descriptions of the previous stages for comparison.

Chall's six stages range from pre-reading at Stage 0 to construction and reconstruction at Stage 5. Stage 0 focuses on pre-reading activities (eg. retelling stories, learning the alphabet, and coloring). Stage 1 includes initial reading and decoding (matching sounds to print). At Stage 2 children begin reading simple stories by using their decoding skills and their ability to recognize vocabulary on sight. At stage 3, readers begin to learn new concepts and acquire knowledge through their reading (usually beginning in about grade 4). Reading material at this stage is usually limited in technical complexity, written in simple language and from one point of view. Stage 3 readers relate print to ideas in order to learn facts. Chall characterizes the reading at the end of level 3 as digest magazines.

At Stage 4, reading is extensive and varied with multiple viewpoints and usually associated with late high school. Material accessible to readers at level 4 are magazines such as *Time* and *Newsweek* and the *New York Times* (Chall, 1985, p.5). Stage 5 is associated with college level and beyond; material at this Stage is complex, abstract and technical. At Stage 5, reading involves the construction and reconstruction of information and the integration of one's own knowledge with that acquired through reading.

Chall (1983) cites Perry (1970) as the greatest influence on the advanced stages of her framework. Perry's seminal work investigated the intellectual development of college-age students through their undergraduate education. As students begin their undergraduate education, they tend to reify knowledge and see their professors as imparters of knowledge and truth. As students learn more, their perception of static truth begins to erode and they recognize the existence of multiple viewpoints (including those of their professors). Subsequently, students accept the co-existence of these many viewpoints and the relative nature of truth. Finally, through a synthesis of the available information, students chose a position to which they are committed.

Thus, Perry's (1970) work indicates that the college years represent a development from the perception of a single point of view to the acceptance of a multiplicity of viewpoints and finally to the formulation of one's own viewpoint. In Chall's model,this is the development from stage 4 to stage 5. The transition from Stage 4 to Stage 5 is the most difficult because readers engage in synthesizing information and critically evaluating conflicting texts. Chall (1983) defines Stage 5 reading as purposeful, strategic, constructive, and highly dependent on prior knowledge. Stage 5 means that "one has the ability to construct knowledge on a high level of abstraction and generality and to create one's own "truth" from the "truth of others" (pg. 24).

The historical changes in the undergraduate student body and curriculum interact substantially with the developmental model. Repeated reading and factual interpretation of text required in the 17th and 18th century at the college level is equivalent to stage 3 within Chall's theory (Heron,p.21). Now college reading demands are much more rigorous and current students, facing many more complex tasks than did their counterparts in the 17th and 18th century, are underprepared for college reading without a stage 4 reading level. At the same time, the diversity of student background and training and the increased accessibility of post-secondary education have made stage 4 reading less likely. Undergraduate instructors must recognize this disparity and be sensitive to the developmental process of learning (and reading specifically) and to the developmental levels at which their students are acquiring knowledge (and interacting with text). Instructional tasks must be aimed at the level of student development while still maintaining academic standards appropriate to higher education. Instructors must gauge a group's ability and not expect students to seek their own truth (Stage 5) when they cannot as yet summarize different points of view (Stage 4).

College-level Reading: Skilled Reading Perspective

Examining skilled reading offers another approach in an attempt to define

college-level reading. I make the assumption here that skilled reading is subsumed under the construct "college reading" since, theoretically, skilled reading could exist outside of the context of college, and therefore not the exact equivalent of college reading. As such, I will first examine a comprehensive definition of skilled reading and then specific reading tasks required in college. In this way, I will focus on how college reading is both skilled and task-specific.

A. Skilled Reading

A blue-ribbon panel of reading theorists and researchers, offered a comprehensive definition of skilled reading in Becoming a Nation of Readers (Anderson, Hiebert, Scott, & Wilkinson, 1985). The authors define skilled reading within five major descriptors:

(1) "Skilled reading is constructive. Becoming a skilled reader requires learning to reason about written material using knowledge from everyday life and from disciplined fields of study."

(2) "Skilled reading is fluent. Becoming a skilled reader depends upon mastering basic processes to the point where they are automatic, so that attention is freed for the analysis of meaning."

(3) "Skilled reading is strategic. Becoming a skilled reader requires learning to control one's reading in relation to one's purpose, the nature of the material, and whether one is comprehending."

(4) "Skilled reading is motivated. Becoming a skilled reader requires learning to sustain attention and learning that written material can be interesting and informative."

(5) "Skilled reading is a lifelong pursuit. Becoming a skilled reader is a matter of continuous practice, development, and refinement (pp.17-18)."

This definition of skilled reading of Anderson *et al.* (1985), much of which mirrors Stage 5 of Chall's Stage Theory (1983), is basic to all reading. The definition addresses using knowledge both from everyday life and from "disciplined fields of study." It underscores the automaticity of basic process as well as metalinguistic knowledge about how to approach texts. It emphasizes the importance of a positive attitude towards reading as well as continual practice to refine the skill.

B. General Skills

It is ironic that almost a century earlier, the Committee of Ten (1894), which served as a precursor to the College Board at the end of the nineteenth century, suggested college reading skills which are not very different from what one would expect today (See Heron, 1989).

- The ability to read for breadth and depth in a given content area
- The ability to comprehend and organize knowledge sufficiently so that one can write clearly and concisely about what one has read
- The ability to read critically in both texts and source material, not only for facts, but to develop a power of comparison and the ability to judge the worth of a book
- The ability to construe evidence from parallel references

These guidelines from the Committee of Ten (1894) assume the skills addressed above in Becoming a Nation of Readers (1985) and focus more directly on how a reader takes knowledge from a text and uses it. These guidelines include the ability to strategize (to read for breadth and depth); to summarize (to write about what one has read); to synthesize (to construe evidence from parallel references) and to evaluate (to judge the worth of a book). These guidelines focus more specifically on the acquisition of knowledge and its use after skilled comprehension takes place. College reading

generally demands a written output of the knowledge one acquires from reading and this additional step marks the distinction between skilled reading and college reading.

The Committee of Ten wrote its guidelines nearly a century ago and yet its emphasis on the reader and the acquisition of knowledge is in keeping with curent reading theory. More recently, the Minnesota Board of Higher Education (1986) attempting to define college level reading focused more specifically on the text and types of comprehension. The panel suggested that reading in college is substantially different from reading in high school: in volume, variety and difficulty. The board offered six broad skill areas necessary for reading at the college level: literal comprehension, text-based inferential comprehension, reader-based inferential comprehension, critical reading, literary understanding and literary appreciation. The panel lamented the lack of systematic reading instruction at the college level suggesting that college reading programs are generally remedial and not used to enhance the reading abilities of "skilled readers." The Board also recommended that high schools implement reading courses focusing on teaching students to (1) use reading as their primary source of information and instruction to (2) read and retain extended expository and narrative texts.

The broad skill areas suggested by the Minnesota Board of Higher Education have some similarities with both the definition of skilled reading and the guidelines from the Committee of Ten although the skills suggested by the Board of Higher Education treat comprehension as an entity rather than as a process of discovery. The Board analyzed comprehension along a continuum from the text itself (summarization) to the reader's interaction with the text (constructive reading), then to the reader's reaction to the text (to write clearly about what one has read) and its critical analysis (to read critically) and finally to the reader's understanding and evaluation of the text in its position as a literary piece (to judge the worth of a book). The skill areas outlined by the Board do not take into account the skills underlying reading

comprehension nor do they specifically address the processes of associating, integrating and synthesizing knowledge acquired through reading as did the Committee of Ten.

In a more general attempt at defining college reading, Cheek (1983) suggests that the construct is not a mere extension of elementary or secondary reading but a unique entity capable of generating its own instructional methodology. He agrees with the Minnesota Board of Education about the importance of college reading instruction and suggests that the instruction should (1) be based on skills that are actually needed for academic tasks, (2) be integrated into the content of the curriculum and (3) make use of peer tutors. Cheek further suggests that "college reading should be a generator of new teaching methods and materials" (p.558). What specific tasks should be included in college reading instruction?

C. Specific Tasks

Defining college-level academic tasks has been of great interest to ESL instructors preparing international students for higher education in the United States (Bridgeman and Carlson, 1984; Horowitz, 1986; Johns, 1981; Kroll, 1979 and Olster, 1980). These researchers have attempted to define the tasks that students face in higher education to provide better pre-university preparation courses. Most of the information has been gathered through surveys, from both students and faculty.

From a survey about the academic needs and the English language proficiency of 133 NNS students at the American Language Institute at the University of Southern California, Olster (1980) reports that the ability to read textbooks was rated as the most important need. Taking notes and asking questions in class was rated next in importance and writing research papers was rated significantly lower.

Both undergraduate and graduate students in Olster's sample agreed on the importance of reading textbooks and taking notes in class, but differences between the two groups emerged. In fact, the responses relating to the reading demands indicated a developmental sequence, from undergraduate to graduate, similar to those found in Perry (1970) and in Chall (1983). First and second year students emphasized the reading of textbooks, graphs, charts and simple lab reports. Juniors and seniors, on the other hand, cited the reading of academic papers and journals and synthesizing research for papers. Tasks reported by the latter were much more similar to those reported by the graduate students.

Johns (1981) surveyed a random sample of faculty at San Diego State University using a very simple questionnaire. Of the four language skills, reading was considered the most important for success in their classes. Listening, writing and speaking followed in order of importance. Johns concluded from the findings that the TOEFL, with its emphasis on reading and listening, may examine those abilities most essential to university success.

In addition to the general language needs cited above, reseachers investigating academic writing tasks also offer insights into college level reading. College writing tasks are generally developed from reading and thus can be examined as a source for reading demands.

Bridgeman and Carlson (1984) gathered survey data from 190 academic departments at 34 universities. Interviews prior to the survey suggested that undergraduate freshmen did most of their writing in English classes and thus the undergraduate surveys were sent only to English departments. Undergraduate writing assignments for freshmen included extensive summaries, research papers, expository and critical writing, and exams with essays. Much less frequent were lab reports, group writing or case studies. The underlying reading skills behind these writing tasks were the abilities to read, summarize, synthesize and evaluate individual texts and multiple

sources. The frequency of these specific types of writing tasks was corroborated by faculty ratings of importance. Moreover, there was great variation in the types of writing tasks and their ratings of importance across graduate departments demonstrating how different academic disciplines demand specific writing (and reading) tasks.

Horowitz (1986) investigated academic writing by requesting recent tests and handouts describing written assignments from faculty members at Western Illinois University. Although the response rate was low (36 faculty members responded), 54 assignment samples emerged.

The types of activities included: synthesis of multiple sources; summary or reaction to a reading; connection of theory and data; case study; research project; report of specific participatory experience; and annotated bibliography. All seven categories of tasks with the exception of "reporting on a specific participatory experience" involved reading. As suggested above in Anderson *et al* (1985), the reading herein reported is constructive, requiring an integration of the various sources of knowledge (eg. lectures, textbooks, class discussions, and other source readings); strategic, involving both extensive and intensive reading, and using skimming and scanning skills; and fluent, requiring the reading of large amounts of material for synthesis.

Summary

In this first section, I have examined "college-level reading" from several different perspectives. First, I discussed some of the historical factors which have impacted on college reading: changes in the curriculum (and in reading) from limited classical and biblical readings for memorization to wide-ranging types of reading in vastly different domains and changes in the student body from the elite of the early years of American University education to those accepted under open admission policies so prevalent today. These two types of changes are in opposition: for just as academic disciplines, fueled by the

technological information explosion, have taken on greater breadth and complexity, the reading ability levels of students, as evidenced by standardized tests and survey data from college faculty and administration, have been declining. Thus, at a time when skilled reading is more essential to college success than ever before, larger numbers of students with less developed reading ability are entering higher education.

Next I examined college reading within a developmental model. In the developmental model, reading is intricately linked to language development, cognition and to learning, itself. At the higher levels of this developmental hierarchy, the ways in which a reader approaches a text serves as a metaphor for how the reader approaches learning. How does the reader create truth from the text? How is the text used to enhance learning? The developmental model of reading offers not only insight to reading goals but also descriptions of levels of reading from which our most enlightened reading evolves.

Finally, I compared a comprehensive definition of skilled reading with general guidelines for college reading and with specific academic tasks reported in survey data. The need for reading skills emerge as primary in the general surveys particularly among the ESL population. Other data reveal that highly complex reading tasks form the basis of most of the academic tasks required in college or university and that these tasks require a level of reading ability at least equal to that of skilled reading as delineated in *Becoming a Nation of Readers*.

What is college-level reading? College-level reading is a dynamic process, in which a reader, armed with prior experience, analyzes and evaluates large quantities of complex and often highly technical textual material written in various styles and representing diverse viewpoints in order to use the information for his or her own purposes. The process of college-level reading also includes automatic processes (eg. decoding unfamiliar words or scanning for specific facts) as well as metacognitive abilities about how to approach a text (eg. skimming for general content, intensive study for retention, or casual reading for pleasure).

SECTION 2: CURRENT DIRECTIONS IN READING ASSESSMENT

The assessment of college-level reading

Given the changes in curriculum and demographics, it can no longer be assumed the college-bound students are prepared to read at the college level. Moreover, tertiary education offers no systematic way of assessing the reading level of students. College reading level, however, is measured, if at all, at two points: prior to admission through scores on the standardized college entrance exams (the SAT and the ACT); and at admission for diagnosis or placement through commercially available reading tests (eg. the Nelson-Denny Reading Test) or by locally designed tests. Both college entrance examinations and commercially available tests for reading make use of the standardized multiple choice test. Standardized multiple choice tests have come under much criticism in recent years (See *Report on the National Commission on Testing and Public Policy* for discussion.) Are standardized, multiple choice reading tests informative enough? Are they in sync with our increasing knowledge about the reading process?

A. *Reading Theory and Assessment*

Reading research over the past decade has greatly enhanced our knowledge about the process of reading. We have moved from a position of text-driven reading where it was the responsibility of the reader to discover the information provided in the text to a more interactive process in which the reader approaches text armed with various types of knowledge: about the specific content, and about its linguistic presentation (eg. vocabulary, sytax and semantics). In this interactive theory of reading (Rumelhart, 1977; Stanovich, 1980; Lesgold and Perfetti, 1981), readers interact with a text making use of their own knowledge and a variety of language skills to give meaning to a text. The interactive theory has also been effectively applied to second language reading (See Carrell, Devine, & Eskey, 1988).

In spite of the fact that current directions in reading emphasize the interactive model, basic reading assessment methods of today are consistent with those of fifty years ago (Farr and Carey, 1986). Johnston (1984) traces the development of reading assessment from the beginning of this century to Thorndike who, because of his prototypical test in 1917, might be considered the "father of modern group tests of reading" (pg.149). Johnston (1984) concludes that the "major approach to the assessment of reading that is currently in use (largely group silent reading tests) seems to have been the result of an ideological thrust that favored the ease of use over all else" (p.153).

Ease is the cornerstone on our reliance on large scale standardized testing in reading as in language testing in general. Given the large numbers of students whose reading ability has to be assessed, standardized tests offer efficient, psychometrically reliable scores for individuals which can normed against scores within a given population. But do standardized reading tests measure a student's ability to read? Many reading researchers say that standardized tests neither measure reading ability nor take into account the knowledge which we have about the reading process (Farr & Carey, 1986; Johnston, 1983; Johnston, 1984; Pearson & Valencia, 1987; Valencia & Pearson, 1987; Valencia, Pearson, Peters & Wixson, 1989; Wixson & Peters, 1987; Wixson, Peters, Weber & Roeber, 1987).

Although some might argue that the internal process of reading has not changed substantially over these fifty years, our empirically-based knowledge about the process has changed considerably (eg. the importance of prior knowledge, the construction of meaning, the interaction of top-down and bottom-up skills). Many reading researchers have called for change in the process of assessment. Curtis and Glazer (1983) conclude that (1) the heterogeneity of test items (on achievement tests in reading) makes it unclear what final scores actually mean and (2) that large-scale standardized tests do not provide the information needed to diagnose reading difficulties. They suggest

the examination of other components of reading (decoding, word access, sentence processing and discourse processing) for a more accurate assessment of reading ability.

B. *Where is large-scale reading assessment headed?*

The call for change in reading assessment has not only been heard, but acted upon at the state and national levels. Two independent groups of reading researchers have initiated major revisions in state-wide assessment in reading in Illinois (Valencia and Pearson, 1987) and in Michigan (Wixson & Peters, 1987). On the national level, the National Assessment for Educational Progress (NAEP) is considering a blueprint for major reforms in its 1992 reading assessment (Rothman, 1990).

The Illinois State Board of Education and the Michigan Department of Education (Valencia *et al.*, 1989), working independently, albeit with an interactive reading model as their common foundation, have developed tests with important similarities. In both tests, the texts are more representative of what students actually read in schools, that is, they are longer and more contextualized (in a format using headings, illustrations, graphs etc. as one might find in a textbook). Both tests view the construction of meaning as a primary test component with three related components: topic familiarity; metacognitive knowledge and strategies; and reading attitudes, habits and self-perceptions" (pg.59).

The Michigan Educational Assessment Program's reading test contains two passages of 500-2,000 words each in two genres: narrative and expository. A pre-test Topic familiarity assessment of 24 test items (12 from each selection) preceeds the reading. Each of the selections is followed by 46 questions: 20 related to the construction of meaning for the selection; 14 related to the student's knowledge about reading (eg. text type, macrostructure, or adjunct

aids); and 12 related to attitudes and self-perceptions as related to each selection. Text mapping and diagrams were used to develop the items in the "Construction of Meaning" section to ensure that only principle ideas were tested in each selection. Student performance on the test is reported in both genres and across all areas of performance as described above.

In the Illinois pilot (Valencia & Pearson, 1987) the format is moving from traditional multiple choice questions to include tasks in which students would (1) evaluate the correctness of multiple acceptable answers; (2) choose from a group of 20 questions 10 which would help a peer best understand the selection; (3) choose the best written summary from a group of several. These pilots also include a measure of prior knowledge and metacognitive judgment (eg. retelling to a different audience).

The changes in the NAEP for 1992 (Rothman, 1990) currently under consideration would include three text types: literature, informative text and documents in complete form rather than a form prepared for the test.

Comprehension would be checked by open-ended questions with written responses and multiple choice questions with more than one appropriate answer. An oral reading measure of one part of the written test is also proposed for some 2,000 4th graders in an effort to examine how an oral reading assessment interacts with a silent reading assessment.

The jury is still out on these new forms of assessment in Illinois and in Michigan and changes in the format of the NAEP are still under consideration. The results from the state pilots need to be examined carefully from both the testing and instructional perspectives. However, these innovations are based on the theoretical foundations of skilled reading.

C. *Skilled Reading and Assessment*

In the first section of this paper, a definition of skilled reading (Anderson, Hiebert, Scott, & Wilkinson, 1985) based on a synthesis of empirical research was discussed. I would like to address each of these components within the framework of reading asssessment.

1. *Skilled reading is constructive*. Skilled readers draw on their own knowledge base to make sense out of text. The importance of what readers bring to the reading task, their background knowledge, has been empirically shown both among native speakers of English (eg. Bransford & Johnson, 1973; Johnson, 1984) and among non-natives (Alderson & Urquhart, 1984b; Hale, 1988; Johnson, 1982). Current standardized tests deny the importance of activating prior knowledge before reading. Short, decontextualized passages are presented without titles, illustrations or purpose to prime the test-takers' schemata (previous experience) about the topic.

Developing a reading test that is completely independent of prior knowledge is an impossibility; it is like trying to develop a reading test devoid of knowledge. No test can account for the variety of experiences and knowledge with which any group of test-takers arrive at the reading test. However, basic knowledge about the content-specific passage information can and should be measured. If not, reading scores may actually be measuring a reader's previous knowledge -a construct quite different from reading as represented by extracting information from a text.

Valencia and Pearson (1987) suggest measuring test-takers' content knowledge, prior to reading, by having them identify ideas which might be included in a selection about a particular topic or by rating the importance of specific vocabulary to the theme of the passage.

2. *Skilled readers are fluent*. They have the ability to use many layers of reading skills simultaneously and automatically. The foundation of fluency is the ability to identify individual words rapidly and accurately in order to understand text. Readers (and even more so test-takers) must "have a [decoding] command that goes beyond simple mastery to automaticity" (Anderson *et al.*, p. 12). Typically, standardized tests do not measure fluency and poor scores may mean poor decoding skills. Lack of speed in word identification detracts from the time needed for comprehension. Poor scores can be misleading since we often have no measure of lack of automaticity but only measures of correctness.

3. *Skilled readers are strategic*. How they read depends on the complexity of the text, familiarity with topic and the purpose. Most tests do not permit skilled readers to use strategies that are important in normal reading, such as the structure which an author provides to organize, learn and ultimately remember the information (Anderson *et al.*, 1985). The passages found on standardized tests are starkly presented without visual or graphic aids to comprehension or any indication of a larger context. Question types tend to be formulaic (Curtis and Glaser, 1983) so that there might be a single, overarching strategy that readers use in this type of reading.

In test taking, readers are devoid of an ecologically valid purpose for reading. The purpose at hand is to answer correctly as many questions as possible, at times, to outwit the test developers rather than to comprehend a passage (Langer, 1987). Strategic reading often becomes strategic test-taking.

4. *Skilled readers are motivated*. I can think of few tasks that are less appealing or more anxiety-ridden than taking a standardized multiple choice test. Moreover, readings on standardized tests are often lifeless in style and content; and focus on literal and sentence-level inferential comprehension items rather than on inference and critical reading (Valencia *et al.*, 1989).

5. *Skilled readers practice, develop and refine their reading over their lifetime.* Standardized reading assessment is a snapshot in the life of a reader in a very controlled reading experience which does little to recognize the variety of reading experiences the test-taker has outside of the testing situation. A portfolio of other reading experiences, interests, and background would offer a more complete assessment of the reader's ability.

Assessment In ESL Reading

ESL reading pedagogy is primarily based on a notion of transfer from L1 (Cummins, 1979). Once enough language is learned in L2, reading strategies from L1 can be applied and successful L2 reading ensues. (See Alderson and Urquhart, 1984a for a comprehensive analysis of the transfer of reading skills across languages). Generally speaking, reading in the ESL classroom is relegated to a secondary position after instruction in syntactic structure, vocabulary and discourse analysis (Allerson and Grabe,1987). Recently, however, the nature of second language reading has attracted a number of researchers, many of whom have examined reading among non-native speakers of English. When examining reading processes among NNS readers, how is reading assessed?

Carrell (1989) provides an overview to second language reading assessment in which she discusses the use of recall of text, both written and oral, and verbal reports to assess L2 reading. She states that these techniques are used with other methods such as multiple-choice questions, cloze and miscue analysis. What do L2 reading researchers use to measure their subjects' reading level or their reading gains?

How do second language reading researchers assess the reading ability of their subjects before and after treatment? For a concrete answer to this question, I examined a decade of empirical research in ESL reading published in the TESOL Quarterly, the principle journal for ESL practitioners. From 1979

to 1989, the TESOL Quarterly contained 16 empirical studies of ESL reading. In these studies reading was assessed 24 times since several studies used multiple measures. The multiple-choice format was used most frequently (11 times). Recall, written (6) and oral (2), was next. True/false and cloze were both used twice and open-ended questions were used once. The three of the four most recent (1988–1989) studies used a multiple-choice format only.

Consequently, it appears from this limited examination of empirical studies, that multiple choice format is the most common form of assessment in ESL, even among ESL reading researchers. Ease of use in scoring, particularly for large-scale testing, and the availability of standardized tests make multiple choice tests the first choice.

L2 reading research has been built on models of reading from L1 and reading assessment has followed suit. Reading tests designed for NNS, such as the Michigan Test of Language Proficiency and the King- Campbell (ELS, 1982), follow the basic pattern of first language reading tests: short reading passages followed by comprehension questions. The TOEFL which assesses reading directly follows the same multiple-choice format. Is multiple-choice testing the most effective measure of reading comprehension? What is the status of the TOEFL as a measure of the language and reading proficiency of undergraduate applicants? How does the TOEFL reflect current directions in reading assessment in L1?

SECTION 3: THE TOEFL AS A READING/LANGUAGE MEASURE

A. What is the TOEFL?

Historical Context

Spolsky (1978) presents a historical description of the development of language testing and thus offers a backdrop for the development of the TOEFL. He divides modern language testing into three distinct eras: the prescientific, the psychometric-structuralist, and the integrative-sociolinguistic. The prescientific was primarily in evidence before the 1920s and was characterized by a lack of concern for statistics, reliance on experienced teachers, who provided passages for translation, free composition, and almost no oral language work. This era preceded the widespread use of psychometric testing.

The psychometric-structuralist trend was based on objective and scientific tests. Assessments developed by teachers were considered unreliable and experts were asked to create objective tests, for the most part limited to reading and listening. Testing during this period involved a marriage of psychology and linguistics. Deriving from the structuralist tradition, test items were "discrete point" reflecting a scientific analysis of language. Some of "the best examples of this type of testing are the standardized language tests developed and administered by the Educational Testing Service: the Graduate Record Exams, the College Board Achievement Tests in foreign languages and the Test of English as a Foreign Language" (Valette, 1967).

Spolsky describes the third era, integrative sociolinguistic, as evolving from two trends in contemporary linguistics: First, language competence was no longer considered merely the sum of discrete elements as had been proposed by the structuralist tradition, but rather as a deep underlying structural system (Chomsky, 1957). Second was the development of sociolinguistics and the enlargement of the notion of language to include the ability to function

within the language, i.e. to perform certain tasks in specific domains within a community of language users (See Young, 1976). Recent advances in language theory have focused on such as ability to use the language and have termed the ability "communicative competence." How to apply the notion of communicative competence to language testing is currenlty being investigated (See Bachman, 1989). For its part, ETS has focused greatly on communicative competence with an internal research project analyzing TOEFL from the communicative perspective (Duran, Canale, Penfield, Stansfield & Liskin-Gasparro, 1987) with an invitational Conference on Communicative Competence and the TOEFL (Stansfield, 1986) and with the inclusion of a writing sample (since 1986) during some TOEFL administrations. It is clear that ETS is attempting to move the TOEFL towards the psychometric-structuralist era from the integrative sociolinguistic.

Development

The TOEFL has become the standard for the measurement of the English language proficiency of non-native speakers of English interested in studying at North American universties or colleges (Alderson, Krahnke and Stansfield, 1987). Developed in 1963 and first administered in 1964, the TOEFL was initially overseen jointly by the College Board and the Educational Testing Service. Since 1973, the TOEFL has been the sole responsibility of the Educational Testing Service. It is estimated that 30 "more than half a million people will sit for the exam during the current testing year" (Educational Testing Service, 1989). The great success of the TOEFL, as discussed in the 1973 (TOEFL) Manual, is due in great part to: its security, availability and high reliability for admission and placement (Oller & Spolsky, 1979).

At the time of the development of the TOEFL, there was a need for language screening for the numbers of international students arriving to study in the United States (Oller and Spolsky, 1979). Therefore, in 1962, the National Council on the Testing of English as a Foreign Language was established to

create such a screening tool. These initial developers of the TOEFL held several major assumptions as cited in Oller and Spolsky (1979): (1) Contrastive analysis (syntactic comparisons between the first language of the test-taker and English) was not used as the basis of the TOEFL; (2) vocabulary level was to be controlled at an unspecified easy level; (3) items were to be written keeping the language as realistic as possible. These early developers created a five-part test, three of which, Vocabulary, Structure and Written Expression, focused on discrete, unrelated questions on specific lexical items, syntax, or writing mechanics. The other two sections, Listening and Reading Comprehension, represented an integrated focus in which several questions related to one passage, conversation or lecture.

Description

ETS moved from the original five-part test and instituted the current three-part in 1976. Section 1, Listening Comprehension, contains three parts. In the first part, examinees listen to a statement and then choose, from four written options, the one which best corresponds to what they have heard. In part 2, examinees hear short conversations followed by questions. Part 3 consists of more formal language (eg. mini-lectures or public announcements) followed by a series of questions.

Section 2, Structure and Written Expression, focuses on grammatical structure and contains one part with incomplete sentences and options and another where sentence errors must be identified.

Section 3, reading comprehension and vocabulary, also consists of two parts: in Part 1, examinees identify word meaning within a sentence context; in Part 2, examinees read short passages followed by comprehension questions. [See Alderson, Krahnke, & Stansfield (1987) and Educational Testing Service (1987, 1990) for more specific information about format.]

Total scores on the TOEFL range from 227 to 677. These total scores and subscores from each of the three sections are reported. 500 is generally considered the minimum score required for undergraduate admission although standards vary considerably across colleges and universities.

As mentioned above, the most recent change in the format of the TOEFL is the Test of Written English. This is a writing sample scored holistically on a six-point scale. Currently, this writing component is included on one-half of the twelve international TOEFL administrations each year.

B. What does the TOEFL measure?

The TOEFL is a measure of language proficiency. Language proficiency, however measured, may be viewed as a tool necessary for success in an English-medium college or university and not as a measure of student aptitude. In general, student aptitude for undergraduate work is measured by the SAT; however, the SAT is not an appropriate measure for NNS applicants because of their generally poor performance on the test, particularly on the SAT-verbal. (See discussion below.) Given the absence of a standardized aptitude measure for NNS applicants, many admissions officers use the TOEFL as such a measure and assume that the TOEFL is predictive of academic success. ETS (1990) continues to maintain that the TOEFL "...is not designed to provide information about scholastic aptitude, motivation, language-learning aptitude or cultural adaptability" (p.15).

The TOEFL has been highly researched within both ETS (ETS, 1989) and the larger academic community which makes use of the exam (Hale, Stansfield, & Duran, 1983) to determine what it measures. The resultant body of research tends to define the TOEFL along a number of parameters: among them, its appropriateness for NNS; comparisons with other measures of language proficiency; its ability to predict academic success; and the unique measurement information from each of the three sections of the test: listening

comprehension, written expression; and vocabulary and reading.

Appropriateness of the test for Non-native speakers of English

The TOEFL is designed for non-native speakers of English. The mean score in a sample of 342,929 NNS students was 509 (ETS, 1990) while average scores for native speakers hover at 630 (Stansfield, 1986). The TOEFL scores of NNS vary greatly permitting easy discrimination among these test-takers. When native English speakers take the TOEFL, their scores skew on the high end resulting in a ceiling effect (Angoff & Sharon, 1971; Clark, 1977; Johnson, 1977), particularly in the Listening Comprehension and Structure Sections. NS show much more variation in the Reading Section.

Comparisons with other measures

a. English language proficiency tests

Not surprisingly, the TOEFL correlates positively with other tests designed to measure English language proficiency among non-native speakers of English: with the Michigan Test of English Language Proficiency [r = .97 and .64] (Cargill-Power, 1981; Dizney, 1965); with the English Language Skills Assessment in a Reading Context [r = .83] (Doherty & Illyin, 1981); with the University of California (Berkeley)/EFL Test [r = .87] (Maxwell, 1966) with the American Language Institute/ Georgetown University Test [r = .84] (AARCAO, 1971). These correlational studies with other English language tests were important for the validation of TOEFL in the early days of its development. However, it is not surprising that these tests, with format and style similar to the TOEFL, correlate with it. High correlations across these language tests suggest the measurement of a similar construct even though this construct may not be language proficiency. Are these tests measuring language proficiency or some other common construct?

b. Aptitude and achievement tests

For NNS applicants, the TOEFL often complements or even supercedes the SAT, the GMAT, or the GRE, standardized tests used for admission purposes. Reliance on TOEFL scores for NNS comes from necessity. NNS generally perform poorly on standardized tests designed for native speakers of English. Angelis, Swinton & Cowell (1979) report mean scores of 274 and 269 for NNS on the verbal sections of the GRE and SAT, respectively, as opposed to mean scores of 514 and 425 for native speakers. Angoff and Sharon (1971) found that NS scored two standard deviations higher on the ACT than their NNS counterparts. SAT-Verbal scores have also been shown to discriminate against Asians born in the United States (Stanley & Abe, 1988). Even NNS 34 Boston high school class valedictorians scored poorly on the SAT-verbal (m=418) (Hartigan, 1990).

Non-native English speakers also score consistently lower on the G-MAT-Verbal than their NS counterparts (Powers, 1980; Wilson, 1982). A synthesis of the findings (Angelis et al., 1979; Wilson, 1982), indicate that only when TOEFL scores reach 625 do scores on verbal aptitude tests "reach the level normally obtained by native speakers" (ETS, 1990, p.29). On the other hand, language proficiency has little effect on quantitative scores (G-MAT and SAT-Math).

In general, comparisons between the TOEFL and verbal tests designed for NS do not adequately measure the ability of NNS. Therefore, the scores of NNS on these aptitude tests should be examined along with their TOEFL scores.

c. Non-standardized measures

Because of its availability and reliability, the TOEFL has also been used as a correlate of several non-standardized measures of language proficiency. Although a discrete item test, until the writing component was added to some adminstrations of the test in 1986, the TOEFL has correlated highly

with theme writing (Carlson, Bridgeman, Camp and Waanders, 1985; Greenberg, 1986; Pitcher and Ra, 1967) and with the number of words in writing samples (Light, 1990). The TOEFL has also been highly correlated to in-depth vocabulary knowledge (Hassan, 1982; Abraham, 1989) and related positively to cloze tests (Ratchford, 1982; and Hinofotis, 1977).

TOEFL has, in essence, positioned itself as such a reliable measure of language proficiency that it often serves as a control i.e. a pre and post measure of language gain in a variety of studies for NNS.

Predictive Ability of the TOEFL

In spite of the caveat from ETS that the TOEFL is a measure of language proficiency and not aptitude, research studies attempting to determine how well the TOEFL predicts academic success have been numerous.

Notwithstanding the variations in defining "academic success," from GPA to number of credits earned or grades in courses in major, these studies generally concur that the TOEFL does not predict academic success to any meaningful extent (Abadzi, 1976; AACRAO., 1971; Andalib, 1976; Bostic, 1981; Graham, 1987; Harvey, 1979; Odunze, 1982; Light, 1990; Light & Massop, 1987; Sokari, 1981; Stover, 1982). Using a measure of English language proficiency to predict academic success is problematic since those students who are admitted into undergraduate study have already reached a threshold level of English. "Proficiency in English of a foreign candidate — at least at the level measured by the TOEFL — is not as stable a characteristic as verbal or mathematical aptitude, rather it is subject to considerable change over relatively short periods of time" (ETS, 1987, p. 27). English language proficiency may be a requirement for academic success but language alone will not ensure academic success. If academic success were contingent on TOEFL scores, then NS students who, on average, score 630 on the TOEFL, should all have high GPAs.

Admission offices seem to expect TOEFL to serve the same purpose as the SAT although the predictive ability of the SAT has been questioned (See Slack & Porter, 1980; Weitzman, 1982). College entrance examination test scores have shown higher relationships to subsequent college grades for native speakers of English than English proficiency test scores for foreign students (Chai & Woelke, 1979). For both native speakers (Wigdor & Garner, 36 1982) and non-native speakers (Andalib, 1976 [r = .33]; Graham, 1984) secondary school grades are better predictors of academic success.

Language learning is clearly a dynamically evolving process and once students have been accepted within an academic environment where language is highly valued, the potential for growth in their language proficiency is immense. In fact, certain studies (Abadzi, 1976; Osanyinbi, 1975) have indicated first year GPA to be more highly correlated with TOEFL scores than that of subsequent years or the cumulative GPA giving evidence for the evolving development of language. Wilcox (1975) corroborated the concept of language development by comparing students with high and low TOEFL scores and finding a threshold at which GPA is no longer related to TOEFL scores. In other words, as a student's English proficiency improves, academic success becomes less related to second language ability (See Graham, 1987, p.515 for discusssion.)

Unique Contributions from all three sections of the TOEFL

A recurring question among language assessment researchers is whether language ability can be captured by a unitary factor (some kind of general factor) or whether it is best analyzed by assessing a number of skills (eg. listening comprehension, speaking, reading and writing as well as syntactic knowledge) (Oller & Hinofotis, 1980; Scholz, Hendricks, & Spurling,1980; Hisama, 1980; Farhady, 1983; Krueger, 1989; Alderson, 1990).

ETS (1990) affirms that all three sections of the TOEFL assess both a common language factor and unique contributions related to that particular skill area. TOEFL scores examined over a two year period, from 1987–1989, showed fairly high correlations among the three subtests. (See Table 1.)

Table 1: Correlations among TOEFL Section Scores (ETS,1990, p. 26)

Section	1	2	3
1. Listening Comprehension	X	.68	.70
2. Structure and Written Expression	.68	X	.79
3. Vocabulary and Reading Comprehension	.70	.79	X

All intercorrelations among the three sections were greater than .68 indicating moderately high correlations. The correlations are not perfect, however, thereby indicating unique contributions from each of the three sections.

Personal experience in interpreting TOEFL scores has suggested that Section 1 (Listening Comprehension) tends to measure the extent to which examinees have interacted with native speakers of English, often specifically how long these students have spent in the United States. Section 2 (Structure and Written Expression) indicates how much classroom study students have had. Section 3 (Vocabulary and Reading Comprehension) seems to measure general academic preparation both in L1 and in English.

In a factor analysis of the subscores, listening comprehension emerged as the principle factor in determining scores across all native language groups (Swinton & Powers, 1982). The power of listening as a factor probably results from a generalized lack of experience among the test-takers in listening to native speakers of English.

Swinton and Powers also indicated differences between structure and written expression (Section 2) and reading and vocabulary (Section 3) between Indo-European language and non-Indo-European groups. For the Indo-European

Reading abilities and TOEFL scores (both total scores and scores for the Reading Section) have been examined in the research through correlational studies, factor analyses and content analysis.

Comparisons between the TOEFL, the Reading Subtest, and various other Tests

Flahive (1980) compared TOEFL scores with the Raven's Progressive Matrices Test (a non-verbal I.Q. test), and three reading tasks: paragraph comprehension from the McGraw-Hill Basic Skills System Reading Tests; a paraphrase recognition task developed by Perkins and Yorio; and a cloze test. Flahive found the Raven's Progressive Matrices Test to be associated with nearly 70% of the variance on the McGraw-Hill Basic Skills Test, underscoring a relation between intelligence and success on reading comprehension tests. Interestingly, the addition of the TOEFL scores to the regression model with the non-verbal I.Q. score added only 1% to the predictive ability. Although not stated in Flahive's (1980) discussion, his findings also could imply that TOEFL scores are related to intelligence.

Others concur with Flahive about a notion of underlying intelligence. Perkins and Pharis (1980) suggest that NNS students have a "deep conceptual ability" (p.146) beneath their surface problems in language and that this ability explains how these students can complete college-level courses successfully in spite of low reading scores on standardized reading tests normed for a NS population. An additional explanation is constant practice in reading in the academic environment. As for the relationships between TOEFL and standardized reading tests, Perkins and Pharis found the following correlations:

Table 1: Correlations between TOEFL and Three Standardized Reading Tests. Perkins and Pharis (1980, p. 145)

GROUP	TOEFL	TOEFL — Reading Subtest
1. Nelson-Denny	.49*	.15
2. Iowa Silent Reading Test	.46*	.23
3. McGraw-Hill Reading Test	X	.91*

*p<.01

The .91 correlation between the McGraw-Hill Reading Test and the TOEFL Reading Subtest in group 3 might be explained by higher language proficiency. The mean TOEFL score for group 3 was 489 while it was 426 and 461, respectively, for groups 1 and 2. Perhaps group 3 had mastered enough English for L1 reading skills to transfer (See Cummins, 1979). It is unfortunate that the complete TOEFL score was not obtained for group 3 for a more direct comparison.

Perkins and Pharis conclude that, based on standardized reading tests, intensive English language students are reading well below "average college freshmen" (p.146) but that the tests may be measuring surface English and not the deep cognitive machinery.

Content of the Reading Subtest of the TOEFL

The reading subtest of the TOEFL in Section 3 follows the typical L1 reading comprehension test and is comprised of five or six short passages followed by several multiple choice questions. According to TOEFL test specifications, students might encounter the reading selections found on TOEFL in books, magazine articles and newspapers (Duran *et al.*, 1987).

In a study comparing the TOEFL scores of NNS and NS, Clark (1977) found that the mistakes made by NS clustered in the Structure and Reading Sections and were "with basic grammar points or straight-forward reading passages of a type college-level students should be expected to handle

without difficulty" (pg.4). Johnson (1977) concurred, finding mistakes made by NS to concentrate on the reading subtest while scores on the listening, structure and vocabulary sections of the TOEFL reached close to maximum.

In an analysis of a specific TOEFL test, Oller (1986) describes the reading passages as "too complex...to have an encyclopedic flavor...ranging from important scientific information to trivia... dry and academic... and missing any sense of disequilibrium, doubt, puzzlement, surprise, or conflict to motivate the text itself or the reading of it" (p.145). He suggests that the reading passages be "... more authentic, relevant, engaging, challenging, and, in the end, more valid" (p.145).

Unlike Oller, Duran *et al.* (1987), in an in depth analysis of the same TOEFL test, found the reading subtest a rich source of language focusing on many grammatical and discourse competencies on topics typical of academic discourse. Duran *et al.* did have some concern over the organization of the reading selections. They found the rhetorical and semantic patterns difficult to code (for analysis) and unlike many of the traditional organizational patterns used in writing.

As for types of comprehension questions in the reading subtest, Duran *et al.* (1988) found them to focus on literal meaning and, in particular, on main ideas and details. Some inferencing of the writer's viewpoint or of information not specifically expressed in the text was also necessary. 50% of the questions were considered unnatural sentence-completion type questions which would not typically be used in real-life communication.

Duran *et al.* summarize that the reading subtest draws on many discrete skills and requires integration "in a manner suggestive of reading in authentic academic settings" (p.42).

Analysis of three TOEFL Tests

To examine more specifically how the TOEFL measures reading ability, the Vocabulary and Reading Comprehension subtests from three retired TOEFL tests were analyzed. All three of these TOEFL tests were retired from use by ETS prior to fall 1989. When ETS decides to retire a test, test administrators permit examinees to leave the examination site with their test booklets. The three tests analyzed in this paper were obtained from students who had used the tests during their final administrations.

a. TOEFL Vocabulary Subtests

The TOEFL vocabulary subtest contains 30 sentences in which there is an underlined word or phrase followed by four options from which the examinee is to choose the one closest in meaning. Ninety such items (30 from each test) were analyzed. The level of difficulty of the 90 words was determined using The Living Word Vocabulary (Dale & O'Rourke, 1981), a graded vocabulary measure. The Living Word Vocabulary scale places a word at a particular grade level after testing its familiarity to students at that grade level. Student familarity is measured by the percentage of students who correctly answer a multiple choice vocabulary question about the word. If between 66% and 85% of the students at a particular grade answered the multiple choice question correctly, then the word was placed at that grade level. If the percentage of correct responses was higher or lower, the word was retested at a more appropriate grade.

Eighty-eight of the 90 vocabulary words on the three TOEFL tests were found on the Living Word Vocabulary Lists. Almost 60% of these words would be familiar to 6 grade students in the U.S. Another 17% would be familiar to most eighth graders. Sixteen percent of the words would be familiar to students in high school and the final 8% (eg. *imprudent* and *heydey*)

familiar to students at the college level. When the grade levels of the vocabulary sections of all three TEOFL tests were compared, the mean grade level of Test 1 (1989) was at the 8th grade while both Tests 2 (1987) and 3 (1989) were at the 6th grade level.

The vocabulary words from these TOEFL subtests were further examined on the Thordike-Lorge (1946) word count list which logs how many times a particular word appears in 1,000,000 words of running text. Sixteen of the 90 words (eg. provide, weight, quarter) were extremely common, receiving the highest score for word count (more than 100 times in 1,000,000 words of text). Another 10 were found more than 50 times; half of the 90 words on these TOEFL subtests were found more than 14 times in 1,000,000 words of text indicating a large percentage of the words on these three tests were quite common. In brief, the vocabulary questions on these three TOEFL tests, in keeping with the premises of the developers of the TOEFL, emphasized vocabulary frequently found in print and familiar to most eighth graders.

b. TOEFL Reading Subtests

The TOEFL reading Subtest includes 5 to 7 reading passages and 30 questions. In this analysis, Test 1 had six passages and Tests 2 and 3, five each. The sixteen passages in these three tests varied in content and readability.

The content of the passages was varied across and within the three subtests. Included were four scientific essays (on whales, magnesium, stars and the development of the central nervous system); three biographies; two philosophical essays (on the future and prejudice); two general interest pieces, similar to press releases on tooth decay and plant watering; one civics essay on the court system; one economics essay on price; and three miscellaneous articles on television, naming practices among the Sioux and the Ordinance of 1784. Test 1 had more selections of general interest. Test 2 had more selections of academic style and Test 3 passages of both types.

Overall readability for the three subtests was determined by a readability analysis of three passages (#1, 3 and 5) from each of the three subtests using the Dale-Chall (1948) readability formula which is based on word familiarity and sentence length. Test 1 had an average grade level of 10 to 11; Test 2, 15 to 16 and Test 3, 11 to 12. The more academic style of the passages on Test 2 were reflected by the higher readability level. Passages on Test 1, with more selection of general interest, were at a much lower readability level.

One drawback in using the Dale-Chall readability formula for non-native speakers surfaced during this analysis. The Dale list of familiar words which is a major part of the readability formula is made up of 3,000 words familiar to children in grade 4. Adults learning a second language may not be familiar with many words commonly known to native-speaking children. For example, the passage on whales in Test 3 included the words: click, squeak, and whistle. These words, although probably common to native speakers, would be very unfamiliar to adult ESL learners.

The 90 comprehension questions in the three subtests were then examined and divided into types. One-half (45) were literal, taken directly from the text. Eleven questions related to identifying the topic or main idea of the passage. Thirteen questions required inferencing from the text and another 13 tested specific vocabulary items. The final 8 questions focused on "text analysis:" pronoun reference within the text and prediction of what would follow in a subsequent paragraph.

The questions were different by test. 60% of the questions in test 1 were literal and found directly in the text and only 6% focused on identifying the main idea. Both the readings themselves and the follow-up comprehension questions were easier on Test 1 than on Test 2 and 3. Easier comprehension questions in Test 1 could be the result of easier passages as shown by the readability. Test 1 was very different from the other two, and possibly written by different authors than Test 2 or 3.

The reading comprehension tends to be the most difficult subtest of the TOEFL as indicated by an average readability level of grades 11-12 for sample passages from the three tests. Thus, TOEFL takers are required to read at a level many NS college students have not yet attained (as discussed in section 1). Yet the vocabulary subtest of the TOEFL is comprised of words familiar to most NS eighth graders. Currently, scores for the vocabulary and reading subtests are reported together as Section 3 by ETS. Reporting both subtests individually would clarify what is being measured: vocabulary knowledge at grade 8 or reading level at grade 11 or 12?

Summary

In this section of this paper, I examined the status of the TOEFL, the most widely used measure of English language proficiency for NNS applicants to undergraduate education. I first described the TOEFL generally, placing it within a global language testing paradigm. I then discussed the development of the TOEFL and described its three sections in depth. Next, I synthesized TOEFL research discussing its appropriateness for NNS, its comparisons with other language measures, its predictive ability, and the importance of all three sections. I then investigated the reading subtest of the TOEFL. I discussed how the TOEFL and the reading subtest correlates with other measures of reading and how researchers described the reading subtest on the TOEFL. I finished with an analysis of the vocabulary and reading subtests of the TOEFL, investigating the level of the vocabulary, the readability of the reading comprehension passages and the types of questions.

Indeed the TOEFL appears to do what it claims to do although many have asked it to do more. It is easily available, accepted world-wide, easily administered and quickly scored. The TOEFL has been highly researched and has attempted to maintain high standards and to incorporate current language proficiency research. In essence, the TOEFL is constrained by the parameters of standardized testing in general, but "given its purpose, examinee

populations and multiple uses, ... TOEFL remains the best of its breed. Beyond those practical limitations that are necessary to its purposes and scope, TOEFL's weaknesses largely reflect the state of the art of language testing (Alderson, Krahnke, & Stansfield, 1987, p. 81)."

ETS has shown interest in improving the measurement of reading on the TOEFL by recently sponsoring an invitational conference to examine the reading and vocabulary subtests of the TOEFL. The current reexamination of large-scale reading assessment (See *The National Commission on Testing and Public Policy, 1990*) and standardized tests in general have prompted this interest from ETS. However, reading on the TOEFL rests squarely on the first language reading assessment paradigm with all its limitations.

Consequently, as on most L1 reading assessments, the reading tasks on the TOEFL do not reflect the natural reading process of skilled reading where background knowledge is activated and reading is purposeful. Nor does the TOEFL require the types of reading skills involved in college-level reading (eg. summarizing, synthesizing, evaluating, etc.). As many have argued, the success of standardized multiple-choice reading tests is that they stand in proxy for "real reading." Change in large-scale assessment is slow and the TOEFL will assess more authentic reading once success in L1 reading measurement is reported.

CONCLUSION

Measuring second language proficiency is a complex procedure especially when admission to higher education is based on that assessment. At present, the TOEFL does an adequate job, and admissions officers have often come to rely on the TOEFL for NNS applicants as they do on the SAT for NS. However, the TOEFL has not been shown, in any practical way, to be predictive of academic success. This finding must be underscored for the many colleges and universities, unfamiliar with the TOEFL, as they look to the international student to increase enrollment.

As for measuring college-level reading, the reading samples from the three TOEFL reading subtests analyzed in this paper were found to be at the late high school and college reading levels with typically academic content. Passages on the tests, similar to those on most standardized reading tests, are presented without a natural context and followed by the usual multiple-choice questions requiring neither the tasks nor skills recognized as important in college-level reading. The TOEFL also offers no unique measure of reading proficiency. The score on the reading subtest is reported along with the vocabulary subtest as a composite section 3 score. The specific section 3 composite score is then reported along with the total score for the test.

Because total scores on TOEFL are more familair to decision makers, a reading vocabulary weakness may be hidden within an acceptable total score. As such, an acceptable total score may reflect the easier, discrete language knowledge as tested on Structure and Written Expression, Section 2, rather than the more complex reading section.

The design of the TOEFL reading subtest is no different from standardized L1 reading tests. And given the numbers and vehemence of researchers and policy makers who are calling for a movement away from traditional multiple-choice testing in L1, L2 reading researchers need to examine these suggestions and decide which ones are appropriate for L2 reading assessment.

RECOMMENDATIONS:

From the number of issues in language testing and assessment, I would like to make the following recommendations for admissions officers, for second language reading researchers and for undergraduate faculty.

Admissions offices should: (1) recognize that TOEFL scores do not predict academic success; (2) continue to rely on TOEFL scores for admission because of the TOEFL's ability to measure language proficiency, a basic tool for academic success; (3) carefully examine all three subscores to assess strengths and weaknesses; (4) track the progress of students accepted at their institutions according to TOEFL scores and (5) keep current on the latest research results from TOEFL studies.

Researchers in second language reading should (1) keep current on the innovations in L1 assessment since reading and assessing language in L2 is based on L1; (2) research new, non-traditional assessment techniques with NNS; (3) delineate more specifically the differences between L1 reading and L2 reading (4) lobby ETS to fund research on the TOEFL reading subtest to better reflect changes in our knowledge about the reading process; and (5) continue to educate instructors about the latest research findings to effect classroom instruction.

Undergraduate faculty who teach NNS students should (1) keep in mind that language inadequacies can be superficial and mask intelligence; (2) recognize the complexity of reading in a second language at the college-level; (3) take into consideration the developmental aspects of both reading and learning for NS and NNS; (4) be aware that simple assignments often involve complex analysis, synthesis and evaluation.

As we strive to develop better language tests, consider the goals suggested by Carroll (1972): that the test measures what we want to know (validity); that it

measure all the varied components of language competence and skills (scope); that the test obtains the best possible measurement within the limits of time and resources (efficiency); and that the test relate to the wider context of instruction.

We would have the perfect English language test for NNS (herein, reading) if we "could put each student into a great variety of absolutely authentic situations in which English must be used to succeed (Barrett, 1990) ...but for now we ...continue to make our testing valid, reliable and practical" (p.8) while we continue to investigate how to make it better.

REFERENCES

Abadzi, J. (1976). Evaluation of Foreign Students Admission Procedures Used at the University of Alabama. *Dissertation Abstracts International, 36*, 7754A. (University Microfilms No. 76-13, 884).

Abraham, P. (1988). *Definition tasks and standardized ESL tests: How do they relate?* Unpublished Paper. Harvard University.

Alderson, J.C. (1990) *The relationship between grammar and reading on an EAP test battery.* Unpublished manuscript. University of Lancaster.

Alderson, J.C., Krahnke, K.J. & Stansfield,C.W. (1987). *Review of English Language Proficiency Tests.* Washington, DC: TESOL.

Alderson, J.C. & Urquhart, A.H. (Eds.) (1984a). *Reading in a Foreign Language.* London: Longham.

Alderson, J.C. & Urquhart, A.H. (1984b). ESP Tests: The problems of students background discipline. In T, Culhane, C, Klein Braley, & D.K. Stevenson (Eds.) *Practice and problems in language testing* (Occasional Papers, No. 29, pp. 1-13). Colchester: University of Essex.

Allerson, S. & Grabe, W. (1986). Reading Assessment in (Eds.) F. Dubin, D.E. Eskey & W. Grabe, *Teaching Second Language Reading for Academic Purposes* (pp.161-181). Reading, MA: Addison-Wesley.

American Association of Collegiate Registrars and Admissions Officers, (AACRAO). (1971). *AACRAO-AID Participant Selection and Placement Study.* Report to the Office of International Training Agency for International Development, U.S. Department of State. Washinton, D.C.: Author.

Andalib, A.A. (1976). The Academic Success of Undergraduate Iranian Students in selected Texas universities. *Dissertation Abstracts International, 36*, 4881A.

Anderson, R, Hiebert, E.H., Scott, J.A., Wilkinson, I.A. (1985). *Becoming a Nation of Readers: The report of the commission of reading*. Washington, D.C.: The National Institute of Education.

Angelis, P.J., Swinton, S.S. & Cowell, W.R. (1979). *The performance of non-native speakers of English on TOEFL and verbal aptitude tests* (TOEFL Research Report, 3). Princeton, NJ: Educational Testing Service.

Angoff, W.H. & Sharon, A.T. (1971). A comparison of scores earned on the Test of English as a Foreign Language by native American college students and foreign applicants to US colleges. *TESOL Quarterly*, 5, 129-136.

Applebee, A.N. (1986). Problems in process approaches: Toward a reconceptualization of process instruction. In A. R. Petrosky & D. Bartholomae (Eds.), *The teaching of writing* (pp.95-113). Chicago: The National Society for the Study of Education.

Bachman, L.F. (1989). Language testing-SLA research interfaces. In R. Kaplan (Ed.), *Annual Review of Applied Linguistics*, 9, 193-209.

Bostic, M.L. (1981). A Correlational Study of Academic Achievement and English as a Second Language (TOEFL). *Dissertation Abstracts International, 43*, 468A (University Microfilms No. 8116851).

Boyer, E.L. (1987). *College: the Undergraduate Experience in America*. New York: Harper & Row.

Brandsford, J.D. & Johnson, M.K. (1973). Consideration of some of the problems of comprehension. In W.G. Chase (Ed.), *Visual Information Processing*. New York: Academic Press.

Bridgeman, B. & Carlson, S.B. (1984). Survey of Academic Writing Tasks. *Written Communication*, 1, 247-280.

Cargill-Power, C. (1981). *A correlational study of TOEFL, Michigan and CELT*. Unpublished Paper. University of South Florida, Tampa.

Carlson, S.B., Bridgeman, B.B., Camp, R. & Waanders, J. (1985). *Relationship of admission test scores to writing performance of native and nonnative speakers of English* (TOEFL Research Report 19). Princeton, NJ: ETS.

Carrell, P.L. (1989). SLA and classroom instruction: reading. In R.Kaplan (Ed.), *Annual Review of Applied Linguistics*. New York: Cambridge University Press.

Carrell, P.L., Devine, J. & Eskey, D. (Eds.). (1988). *Interactive Approaches to Second Language Reading*. New York: Cambridge University Press.

Carroll, J.B. (1972). Fundamental Considerations in Testing the English Proficiency of Foreign Students. In H.B. Allen & R.N. Campbell (Eds.) *Teaching English as a Second Language: A Book of Readings*. NY: McGraw-Hill.

Carroll, J.B. (1977). Developmental Parameters of Reading Comprehension. In J.T. Guthrie (Ed.), *Cognition, curriculum and comprehension*. Newark, DE: International Reading Association.

Chai, S.H. & Woelke, P.L. (1979). The predictive ability of standardized tests of English as a foreign language. In R. Silversetein (Ed.), *Proceedings of the Third International Conference on Fronties in Language proficiency and Dominance Testing. Occasional Papers on Linguitics, No. 6*. Carbondale: Southern Illinois University. (ERIC Document Reproduciton Service No. ED 186 476).

Chall, J.S. (1983). *Stages of Reading Development*. New York: McGraw-Hill.

Chall, J.S. (1985). *The reading of community college students: How well are they doing? How well should they be doing?* Paper presented at Riverside Community College, Riverside, CA.

Chase C.I. & Stallings, W.M. (1966). Tests of English language as predictors of success for foreign students. *Indiana Studies in Prediction, 8*. Bloomington: Bureau of Educational Studies & Testing, Indiana University.

Cheek, D.H. (1983). The evolving definition of college reading. *Journal of Reading*, 26, 556-558.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Clark, J.L.D. (1977). *The Performance of Native Speakers on the Test of English as a Foreign Language*. (TOEFL Research Report,1) Princeton, NJ: Educational Testing Service [ED218 920].

Committee of Ten. (National Educational Association of the United States). (1894). *Report of the committee of ten on secondary school studies; with the report of the conference arranged by the committee*. New York: American Book Company.

Cross, K.P. (1976). *Accent on Learning*. San Fransicso: Jossey-Bass.

Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism*, 19, 197-205.

Curtis, M.E. & Glaser, R. (1983). Reading Theory and the Assessment of Reading Achievement. *Journal of Educational Measurement*, 20(2), 133-147.

Dale, E. & O'Rourke, J. (1981). *The living word vocabulary: a national vocabulary inventory: the words we know.* Elgin, IL: Field Enterprises Educational Corporation.

Dale, E. & Chall, J.S. (1948). *A formula for predicting readability.* Bureau of Educational Research: Ohio State University.

Dizney, H. (1965). Concurrent Validity of the Test of English as a Foreign Language for a Group of Foreign students at an American University. *Educational and Psychological Measurement*, 25, 1129-31.

Doherty, C. & Illyin, D. (1981). *Technical Manual for ELSA: English Language Skills Assessment in a Reading Context*. Rowley, MA: Newbury House.

Duran, R.P. (1988) Validity and language skills assessment: Non-English background students. In H. Wainer & H.I. Braun (Eds), *Test validity* (pp. 105-127). Hillsdale, NJ: Erlbaum.

Duran, R.P., Canale, M., Penfield, J., Standsfield, C.W., & Liskin-Gasparro, J.E. (1987). TOEFL from a Communicative Viewpoint on Language Proficiency: A working paper, In R. Freedle & R.P. Duran (Eds.) *Cognitive and Linguistic Analysis of Test Performance* (pp. 225-244). Norwood, NJ: Ablex.

Educational Testing Service. (1987). *TOEFL Test and Score Manual*. Princeton, NJ: Educational Testing Service.

Educational Testing Service. (1989). *TOEFL Test Score and Manual*. Princeton, NJ: Educational Testing Service.

Educational Testing Service. (1990). *TOEFL Test Score and Manual*. Princeton, NJ: Educational Testing Service.

ELS (English Language Services). (1982). *An English Reading Test*. (Harold King and Russell Campbell).

Eurich, A.C & Kraetsch, G.A. (1981). A 50-year comparison of University of Minnesota freshman's Reading Performance. *Journal of Educational Psychology*, 74, 660-665.

Farr, R. & Carey, R.F. 1986. *Reading: What Can Be Measured?* (Second Edition). Newark, DE: International Reading Association, Inc.

Farhady, H. (1983). On the plausibility of the unitary language proficiency factor. In J.W. Oller, Jr. (Ed.), *Issues in language testing research* (pp. 11-28). Rowley, MA: Newbury House.

Flahive, D. E. (1980). Separating the g factor from reading comprehension. In J.W. Oller & K. Perkins (Eds.) *Research in Language Testing* (pp.34-46). Rowley, MA.: Newbury House.

Flower, L. & Hayes, J.R.(1981). A cognitive process theory of writing. *College Composition and Communication*, 32, 365-387.

Fulwiler, T. (1984). How well does writing across the curriculum work? *College English*, 46, 113-125.

Graham, J.C. (1984). *Predictors of ESL student academic success*. Unpublished manuscript. University of Maryland, Baltimore County.

Graham, J.C. (1987). English language proficiency and the prediction of academic success. *TESOL Quarterly*, 21, 505-521.

Greenberg, K.L. (1986). The development and validation of the TOEFL writing test: A discussion of TOEFL Research Reports 15 and 19. *TESOL Quarterly*, 20 (3), 531-544.

Hale, G.A. (1988). Student major field and test content: Interactive effects on reading comprehension in TOEFL. *Language Testing*, 5, 49-61.

Hale, G.A., Stansfield, C.W. & Duran, R.P. (1983). *Summaries of TOEFL Studies, 1963-1982*. (TOEFL Research Report, 16). Princeton, NJ: Educational Testing Service.

Hartigan, P. (1990, June 7). Against all odds. *The Boston Globe*, pp. 77, 83.

Harvey, M.J. (1979). *Academic achievement as predicted by the Test of English as a Foreign Language*. Unpublished master's thesis, Portland State Universty.

Hassan, K.I. (1982). *The correlation between performance on the WAIS-R Vocabulary subtest and the TOEFL*. Unpublished masters's thesis, Washington University, St. Louis, MO.

Heron, E.B. (1989). *The Dilemma of College Reading Instruction: A Developmental Analysis*. Doctoral Dissertation. Harvard University: Cambridge, MA.

Hillocks, G. Jr. (1986). *Research on Written Compostion: New Directions for Teaching*. Urbana, IL: National Conference on Research in English/ERIC Clearinghouse on Reading and Communication Skills.

Hills, J.R. (1978) Review of the ACT assessment program in O.K. Buros (Ed.), *The mental measurement yearbook*. Highland Park, NJ: Gryphon Press.

Hinofotis, F.B. (1977). An Investigation of the concurrent validity of cloze testing as a measure of overall proficiency in English as a second language. *Disseration Abstracts International, 38*, 762A-63A.

Hisama, K. K. (1980). An analysis of various ESL proficiency tests. In J. W. Oller & K. Perkins (Eds.), *Research in Language Testing* (pp.47-53). Rowley, MA: Newbury House.

Hofstadter, R. and Smith, W. (1961.) *American higher education: a documentary history*. Vs. 1 & 2. Chicago: University of Chicago Press.

Horowitz, D. 1986. What Professors Actually Require: Academic Tasks for ESL. *TESOL Quarterly*, 20, (3), 445-461.

Inhelder, B. & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Johns, A.M. (1981). Necessary English: A Faculty Survey. *TESOL Quarterly*, 15, 51-57.

Johnson, D.C. (1977). The TOEFL and domestic students: conclusively inappropriate. *TESOL Quarterly*, 11, 79- 86.

Johnston, P.H. (1983). *Reading Comprehension Assessment: A Cognitive Basis*. Newark, Delaware: International Reading Association.

Johnston, P.H. (1984). Assessment in Reading. In P.D. Pearson (Ed.), *Handbook of Reading Research*. New York: Longman.

Kretschmer, J.C. (1972). *Measuring rea... g comprehension*. Ann Arbor, Michigan: University Microfilms International, (72-27, 045).

Kroll, B. (1979). A survey of writing needs of foreign and American college freshman. *ELT Journal*, 33, 219-226.

Krueger, P. (1989). *Identification of the relationship between the measure of English language proficiency and academic success among foreign students at one American university*. Doctoral Dissertation. Northeastern University.

Langer, J.A. (1986). The construction of meaning and the assessment of comprehension: an analysis of reader performance in standardized tests Items. In R. Freedle and R.P. Duran (Eds.) *Cognitive and Linguistic Analysis of Test Performance* (pp. 225-244). Norwood, NJ: Ablex.

Lederman, M., Ribaudo, M. & Ryzewic, D. (1985). Basic skills of entering college freshmen. A National Survey of Policies and Perceptions. *Journal of Developmental Education*, 9, 10-13.

Leedy, P. (1959). *A history of the origins and development of instruction in reading improvement at the college level*. Unpublished Doctoral Dissertation, New York University.

Lesgold, A. & Perfetti, C. (Eds.) (1981). *Interactive Processes in Reading*. Hillsdale, N.J.: Erlbaum.

Light, R.L., Xu, M. & Mossop, J. (1987). English Proficiency and Academic Performance of international students. *TESOL Quarterly*, 21, 251-261.

Light, R.L. & Ten-yuan, W. (1990, March). *Soviet Students at U.S. College: Social Perceptions, Language Proficiency, and Academic Success*. Revised version of a paper presented at the TESOL Conference, San Francisco, CA.

Maxwell, A. (1966). *A comparison of TOEFL and the UCB/EFL Test*. Unpublished master's thesis. Sacramento State College, CA.

Maxwell, M. (1979). *Improving student learning skills*. San Francisco: Jossey-Bass.

Minnesota Higher Education Coordinating Board. (1986). *Defining college-level reading skills. Report of the Task Force on Definition of College-level skills*. [ED278 303].

National Commission on Testing and Public Policy. (1990). *From gatekeeper to gateway: transforming testing in America*. Report of the National Commission on Testing and Public Policy. Boston College, Chestnut Hill, MA: Author.

Oakley, F. (1990, March). Despite its critics, undergraduate education is a success. *Chronicle of Higher Education*, 36,(26).

Odunze, O.J. (1982). Test of English as a foreign language and first year GPA of Nigerian students. *Dissertation Abstracts International*, *42*, 3419A-20A.

Oller, J.W. Jr. (1986). Communication theory and testing: what and how. In C.W. Stansfield (Ed.), *Toward communicative competence testing: Proceedings of the second TOEFL invitational conference* (Research Report No. 21) (pp. 104-155). Princeton: Educational Testing Service.

Oller, J.W. Jr. & Hinofitis, F.B. (1980). Two mutually exclusive hypotheses about second language ability: indivisible or partially divisible competence. In J.W. Oller & K. Perkins (Eds.) *Research in Language Testing*. (pp.13-23). Rowley, MA: Newbury House.

Oller, J.W. Jr. & Spolsky, B. (1979). Test of English as a Foreign Language, Some Major Tests. In B.Spolsky (Ed.), *Advances in Language Testing Series, 1*. Arlington, VA: Center for Applied Linguistics, (pp.92-100) [ED 183 004].

Olster, S. 1980. A survey of the academic needs for advanced ESL. *TESOL Quarterly*, 14, 489-502.

Osanyinbi, J.A. (1975). A concurrent validity study of the West African School Certificate and General Certificate of Education English Language Examination using Educational Testing Services's Test of English as a Foreign Language as the criterion measure. *Dissertation Abstracts International, 35*, 5130A-31A.

Pearson, P.D. & Valencia, S. (1987). Assessment, accountability, and professional perogative. In Readance, J.E. & Baldwin, R.S. (Eds.) *Research in Literacy: Merging Perspectives*. 36th Yearbook of the National Reading Conference. Rochester, NY: National Reading Conference.

Perkins, K. (1988). Measuring ESL readers' ability to apply reasoning in reading: a validity study of the TOEFL reading subtest. *Journal of Research in Reading*, 11, 36-49.

Perkins, K. & Pharis, K. (1983). TOEFL score in relation to standardized reading tests. *Research in Language Testing*. Ed. J.W. Oller, Jr. & K. Perkins. (pp. 142-146). Rowley, MA: Newbury House.

Perry, W. (1970). *Forms of Intellectual and ethical development in the college years: A scheme*. New York: Holt, Winehart & Winston.

Piaget, J. (1970). *Structuralism*. New York: Basic Books.

Platt, G. (1987). Should college teach below college-level courses? *Community College Review*, 14, 19-25.

Pitcher, B. & Ra, J.B. (1967). *The relation between scores on the Test of English as a Foreign Language and ratings of actual theme writing* (Statistical Report, 67-9). Princeton, NJ: ETS.

Plisko, V.W. & Stern, J.D.(Eds.). (1985). *The condition of education, 1985*. National Center for Educational Statistics. Washington,D.C.: US Department of Education.

Powers, D.E. (1980). *The Relationship between scores on the Test of Graduate Management Admissions Test and the Test of English as a Foreign Language (TOEFL Research Report, 5)*. Princeton, NJ: Educational Testing Service [ED 218 329].

Raimes, A. (1985). What Unskilled ESL Students do as they write: A Classroom study of composing. *TESOL Quarterly*, 19, 229-258.

Ratchford, D.L. (1982). Reading Ability and entering freshmen international students at a Southwestern State University: Some implications. *Dissertation Abstracts International, 42*, 3088A-89A.

Rothman, Robert. (1990, February, 14). Aiming for a 'definition of literacy,' NAEP considers the 1992 Reading Test. *Education Week*, IX, (21), pp. 1, 22.

Rumelhart, D. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and Performance*, VI, 573-603. New York: Academic Press.

Scholz, G., Hendricks, D., Spurling, R., Johnson, M. & Vandenberg, L. (1980). Is language ability divisible or unitary? A factor analysis of 22 English language proficiency tests. In J. Oller & K. Perkins (Eds.), *Research in Language Testing* (pp.24-33). Rowley, MA: Newbury House.

Silber, J. (1987). The paradox of democratic education: the American experience. *Higher Education in Europe*, 12, 38-42.

Slack, W.V. & Porter, D. (1980). The Scholastic Aptitude Test: A Critical Appraisal. *Harvard Educational Review*, 50, 154-175.

Sokari, H. (1981). Predictors of college success among foreign students from various ethnocultural backgrounds. *Dissertation Abstracts International*, 41, 3543A-44A.

Spack, R. & Sadow, C. (1983). Student-teacher working journals in ESL composition. *TESOL Quarterly*, 19, 703-725.

Spack, R. (1988). Initiating ESL Students into the Academic Discourse Community: How Far Should We Go? *TESOL Quarterly*, 22, 29-51.

Spolsky, B. (1978). Introduction: Linguisitics and Language Testers. In B. Spolsky, (Ed.), *Advances in Language Testing Series, 2* (pp. v-x). Arlington, VA: Center for Applied Linguistics.

Stanley S. & Abe, J. (1988). *Predictors of academic achievement among Asian Americans and white students*. (College Board Report, No. 88-11.) [ED 303 555].

Stanovich, K.E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16, 32-71.

Stansfield, W. (Ed.) (1986). *Toward Communicative Competence Testing: Proceedings of the Second TOEFL Invitational Conference*. (TOEFL Report, 21) Princeton, NJ: ETS.

Stover, A.D. (1982). Effects of language admission criteria on academic performance of non-native English-speaking students. *Dissertation Abstracts International, 42*, 4374A-75A.

Swinton, S.S. & Powers, D.E. (1982). *Factor analysis of the Test of English as a Foreign Language for several language groups*. (TOEFL Reseach Report, 6; ETS Research Report, 80-32). Princeton,NJ: ETS.

Thorkdike, E.L. & Lorge, I. (1946). *The Teacher's Word Book of 30,000 Words*. Bureau of Publications, Teacher's College, Columbia University.

Valencia, S. & Pearson, P.D. (1987). Reading assessment: time for a change. *Reading Teacher*, 40, 726-32.

Valencia, S., Pearson, P.D., Peters, C. & Wixson, K. (1989). Theory and Practice in Statewide reading assessment: closing the gap. *Educational Leadership*, 46, 57-63.

Vallette, R. (1977). *Modern Language Testing*. New York. Harcourt, Brace, Jovanovich.

Weitzman, R.A. (1982). The prediction of college achievement by the Scholastic Aptitude Test and the high school record, *Journal of Educational Measurement*, 19, 179-191.

Wigdor, A.K. & Garner, W.R. (Eds.) (1982). *Ability testing: Uses, consequences and controversies (Part II)*. Washington, DC: National Academy.

Wilson, K.M. 1982. *A comparative analysis of TOEFL examinee characteristics, 1977-1979*. (TOEFL Report, 11). Princeton, NJ: ETS.

Wilcox, L.O. (1975). The prediction of academic success of undergraduate university students. *Dissertation Abstracts International*, 35, 6084B.

Wixson, K.K., Peters, C.W., Weber, E.M. & Roeber, E.D. (1987). New directions in statewide reading assessment. *Reading Teacher*, 40, 749-754.

Wixson, K, K. & Peters, C.W. (1987). Comprehension assessment: implementing an interactive view of reading. *Educational Psychologist*, 22. 333-356.

Young, L.A. (1976). (Ed.), *Oral language tests for bilingual students: an evaluation of language dominance and proficiency instruments*. Portland, OR: Northwest Regional Educational Laboratory.

Zamel, V. 1982. The Process of Discovering Meaning. *TESOL Quarterly*, 16 (2), 195-210

Zamel, V. 1983. The Composing Process of Advanced ESL Students: Six Case Studies. *TESOL Quarterly*, 17(2), 165-188.

Zikopoulos, M. 1990. *Open Doors: 1989-1990, Report on International Educational Exchange*. New York: Institute of International Education.